A Proposed Model for Improving the Performance of Knowledge Bases in Real-World Applications by Extracting Semantic Information

Abdelrahman Elsharif Karrar
College of Computer Science and Engineering
Taibah University, Medina, Saudi Arabia

***Abstract*—Knowledge Bases are information resources that convert factual knowledge to machine-readable formats to allow users to extract their desired data from multiple sources. The objective of knowledge base population frameworks is to extend KBs with semantic information to solve fundamental artificial intelligence problems such as understanding human knowledge. Information extraction entails the discovery of critical knowledge facts from unstructured text, which is important in the population of knowledge bases. The objective of this paper is to explore the concept of information extraction as a technique for accelerating the performance of knowledge bases with minimal annotation efforts for real-world applications such as content recommendation during a web search. This entails performing slot filling operations for data collection from large KBs and applying probabilistic estimations to determine the accuracy of the new information. The results are then used to explore the feasibility of applying knowledge bases to real-world tasks such as user-centric information access by encoding entities with deep semantic knowledge.**

Keywords—Semantic information extraction; knowledge base; slot filling; content recommendation

I. Introduction

Knowledge Base (KB) refers to a specially designed resource for gathering and processing knowledge in logical statement formats that define the relationship between graphical entities. Knowledge Bases utilize a relational knowledge representation framework implemented on artificial intelligence, logic, and semantic networks [1]. Facts representation through KBs follows the guidelines by Resource Description Framework (RDF) in the definition of variable relationships among entities, predicates, and values forming triples such that entities represent people or objects, predicates define entity relationship, and values represent other entities, types, attributes, and values [2]. Triples represent existing facts as illustrated in Table I.

Triples in a knowledge base can be aggregated into a graph composed of directed edges representing relationships and nodes representing values and entities. Edge directions reflect the subject entities in specific triples in the condition of two entities. This implies that edges bridge subject entity to object entity. Different edge types are used to represent various relations through structures known as Knowledge graphs, which enhance the visualization and comprehension of KG structures. [3]

DBpedia is an example of a Knowledge Database, which has been developed by research communities to provide an effective framework for knowledge representation as shown in Fig. 1 [4] [5].

Technology companies such as Google, Microsoft, Facebook, and Yahoo construct and manage in-house Knowledge Bases to perform functions such as answering questions and data querying. The most common knowledge bases operated by technology companies include the Microsoft Graph Satori, Facebook Entity Graph, and Yahoo Knowledge graph illustrated in Table II alongside their relation types, number of entities, and the volume of facts [6].

TABLE I. INSTANCES OF TRIPLES IN KNOWLEDGE BASE

Entity	**Predicate**	**Value**
Donald Trump	Age	75
Donald Trump	Profession	Politician, Actor
Donald Trump	Starred in	Apprentice TV show
Apprentice	Genre	Reality Competition
Apprentice	Release Date	January 2004

TABLE II. FEATURES AND ATTRIBUTES OF POPULAR SCHEMATIC KNOWLEDGE BASES

Knowledge Base	**Entities**	**Relation Types**	**Facts**
Google Knowledge Graph	570 million	35,000	18 billion
Yahoo Knowledge Graph	3.4 million	800	1.391 billion
Freebase	40 million	35,000	637 million
DBpedia	4.6 million	1,367	539 million
YAGO2	9.8 million	114	447 million
Wikidata	18 million	1,632	66 million

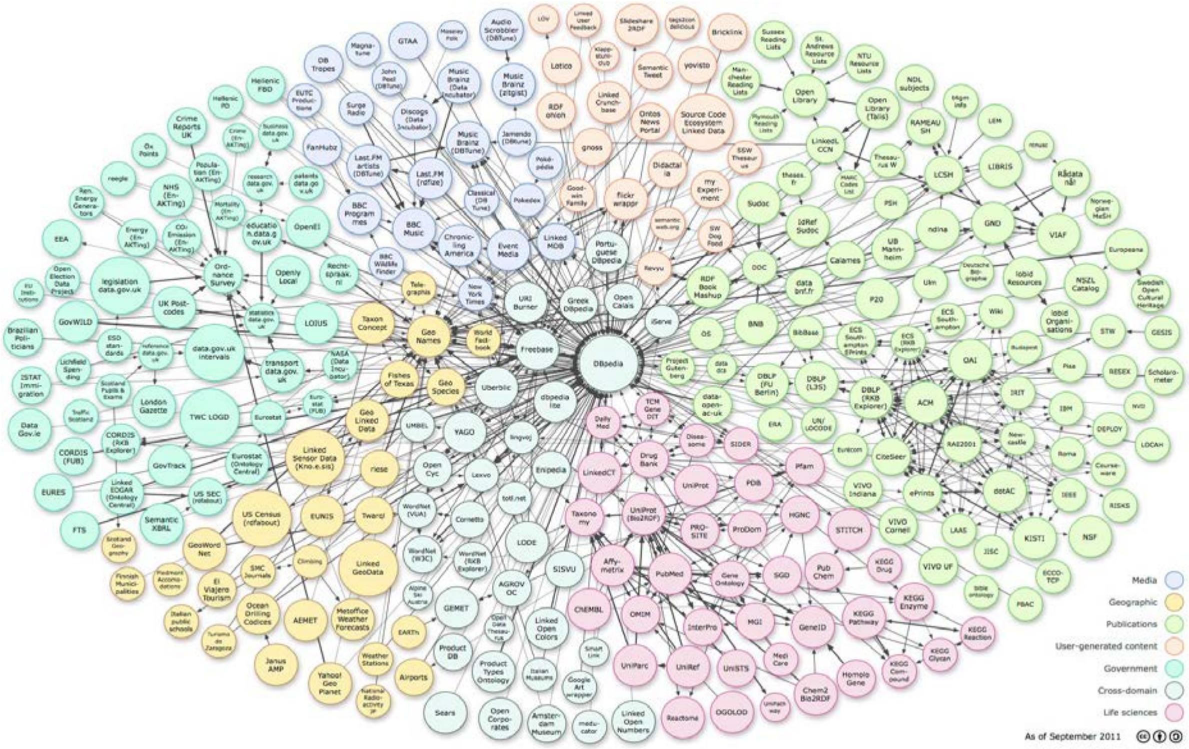

Fig. 1. The Structure of DBpedia Knowledge Base.

Knowledge bases differ from traditional databases in their approach to information management since they are focused on “tables” and “records”, which make them efficient when the discovery of new information is not a priority [7]. Knowledge bases are particularly important in domains where the flexibility to link multiple types of information is required. Some of the unique advantages of knowledge bases over traditional data warehouses include:

- Entity-centric: All data is stored based on entity relevance.
- Schema-less: There are no prior requirements for a schema in the knowledge structure.
- Metadata Rich: This contains self-describing metadata streams, which can be easily scaled and integrated across multiple domains.

A. Applications of Knowledge Bases

Knowledge Bases allow for the semantic structuring of computer-readable information, which is a valuable requirement in the construction of intelligent systems [8]. Knowledge bases are a source of power to various big data applications in multiple scientific and commercial domains such as the integration into Google search engine, which stores approximately 0.57 billion entities and 18 billion facts [9]. The Google Knowledge Graph plays an important role in the identification and disambiguation of textual entities to generate enriched search results by semantic structuring of summaries while providing links to related content during explanatory search [10]. Companies typically rely on knowledge bases in gathering information about various entities and their relationships for optimal reuse efficiency in a domain. Knowledge bases are typically used in querying and displaying entity information, recognizing and extracting context, linking entities to data sources and content, discovering and suggesting related information, semantic parsing, and answering questions in technology platforms such as social media and AI-driven virtual assistants.

The role of knowledge bases in utilizing semantic information generated from knowledge graphs to enrich search results is an important milestone towards the transformation of text-based search engines such as Google into semantically-aware question answering platforms. The concept of knowledge graphs has been prominently demonstrated in Watson; a question-answering platform developed by IBM. Watson used a combination of information sources including Freebase, DBpedia, and YAGO to win the game of Jeopardy against a team of human experts [11]. Structured knowledge repositories are integrated into digital assistants such as Amazon Echo, MS Cortana, and Siri by Apple. Knowledge bases such as Freebase store general data generated by its community members from multiple sources including wiki contributions. Knowledge bases have been applied in the Internet Movie Database (IMDb), which is an online storage platform for information related to video games, television programs, and films including character biographies, reviews, crew information, and plot summaries [12].

B. The Concept of Information Extraction

Information Extraction (IE) refers to a process through which structured data is generated from semi-structured or unstructured machine-readable formats [13]. The traditional information extraction systems are used for the efficient extraction of data from isolated documents using advanced information retrieval methods for data scattered in multiple documents. The systems are capable of identifying the documents containing relevant information and extracting specific facts concerning entities that are conflicting, complementary, or redundant [14].

The first step of data gathering in IE systems is consolidating the known information regarding a specific query entity then searching multiple sources for related information. For instance, if a query 'Donald Trump' is made on an IE system, the objective of slot-filling components is to consolidate information on Donald Trump's place and date of birth, occupation, marital status, education, and any other pre-defined attribute through a process known as 'filling' then adding other related information as recommendations [15]. This process is known as relation extraction since it entails classifying related entities to a relation of interest. For instance, if the system reads a statement 'Donald Trump was born in New York City, the relation born in is extracted to generate search results as (Donald Trump, New York). Information extraction systems are designed to automatically filter information from a pool of sources to fill the missing knowledge base attributes through slot filling before the entities are liked based on their relations.

This research aims at developing a model for improving knowledge basis by extracting information by answering the following research questions;

1) What techniques can be used to construct knowledge bases?

2) How can the accuracy of information extracted from knowledge bases be extracted?

3) In what ways can the efficiency of knowledge bases be improved to perform other tasks such as content recommendation?

This research paper is organized in sections including a review of published literature on the use of knowledge graphs in spoken language understanding, confidence estimation of extracting information systems and the effectiveness of information extraction techniques in improving natural language processing to enrich annotations as well as its role in content recommendation by user profiling in Section II, Section III focuses on the implementation of information extraction techniques and models for improving knowledge bases based on the spoken language understanding (SLU) framework, Section IV explores a high-performance content recommendation model for efficient information extraction from knowledge bases. Section V of this research paper discusses conclusions based on the experimental results and Finally, Section VI provides recommendations for future studies.

II. Literature Review

A. *The Population of Knowledge Graphs in Spoken Language Understanding (SLU)*

The role of SLU techniques in knowledge bases is to perform slot filling tasks and user intent determination, especially in call routing systems, which are integrated with utterance classification capabilities whereby a speech utterance S_i is categorized into one of M semantic categories, $\hat{C}_r \in C = \{C_{1\ldots,}C_M\}$ given that r represents the utterance index [16]. Researchers have recently developed an advanced slot filling method that involves framing tasks in the form of sequence classification problems to identify the phrase boundaries and labels in a semantic template through deep learning [17] [18]. Slot filling tasks in SLU are defined in the Knowledge Base Population (KBP) whose objective is consolidating information from a large multisource corpus for specific attributes of a query entity. Knowledge graphs are powerful and valuable tools for simplifying research tasks such as computing entity weights to allow the allocation of probabilistic weights in the process of enriching semantic knowledge when detecting SLU relations [19] [20] proposed advanced techniques for processing search queries through semantic parsing in multi-turn dialog systems based on unsupervised natural language processing models.

B. *Confidence Estimation in IE Systems*

According to [21] confidence estimation refers to a machine learning technique that is used to estimate the confidence scores of a specific output in applications such as machine translation and semi-supervised extraction of relations. The confidence scores of output from speech recognition machines can be computed using a maximum entropy model as described by White and Markov models for singleton tokens.

Another research paper [22] proposed an efficient confidence estimation approach for IE outputs based on machine learning models. This approach worked by computing confidence scores for both multi-field records and extracted fields based on the linear-chain Conditional Random Field (CRF) framework.

However, the machine learning approach is simpler compared to the slot filling technique, which performs complex tasks such as sophisticated inference and co-reference resolution across multiple documents [23]. Inaccurate values extracted in the slot filling operations for KBPs in multiple systems are filtered using techniques such as weighted voting, unsupervised multidimensional truth-finding, heuristic rules, and supervised learning [24].

C. *Rich Annotations*

Natural Language Processing (NLP) operations such as extracting information can be improved by leveraging user reviews to customize a system to perform personalized searches [25]. Since user reviews may not be readily available, labels created by human annotators, which apply to a range of supervised learning methods can be used to customize the information retrieval system as proposed by [26]. In this case, the traditional machine learning paradigm may be incorporated with a privileged knowledge model to enable the system to accommodate more annotator labels. Recent studies observe an issue with the underutilization of human annotators due to the inclusion of rich annotations into various classification problems [27] [28].

The approach to learning new information through error corrections is conceptualized from the Transformation-based

Error-Driven learning that has been applied to a range of natural language processing operations such as word sense disambiguation, part-of-speech tagging, and semantic role labeling [29]. Rules of transformation in these error-correction techniques are learned automatically based on iteration contexts in each sentence.

D. *Content Recommendation by user Profiling*

The effectiveness of information extraction techniques for improving Knowledge [30] Bases may be improved through user profiling using factorizing machines and recommendation systems. Research studies suggest that the primary objective of user profiling in IE systems is to align user interests with the recommended items for example in online shopping platforms such as Amazon, the content recommendation in Netflix, or web search customization for enhanced user experience in Google [31].

The functional mechanism of recommendation systems in data extraction may be through content-based recommendation or collaborative filtering, which utilizes matrix factorization and nearest neighborhood techniques to compute user collaboration scores [32].

However, content-based recommendation algorithms work by extracting the unique and dominant attributes that explicitly link users to items, especially in systems with multiple cold start items [33].

According to [34] and [35] researchers have proposed improved approaches for content recommendation by user profiling based on activity ranking, hypergraph learning, latent factor models, probabilistic models, and spatial-temporal model.

A study by [36] observes that developers are now more focused on embedding recommendation and user profiling systems with hierarchical knowledge repositories for the creation of personalized entity recommendations based on knowledge and user activity log obtained from freebase.

A content-based recommendation model proposed by [37] implements a spreading activation algorithm on the DBpedia categorization structure to extract [38]information on user preferences and interests. This technique was later applied to music entity recommendation by Linked Data Semantic Distance (LDSD) with DBpedia by [39] and to movie recommendation by [40]. The recommendation systems are capable of modeling user preferences by exacting information from multiple sources such as implicit and explicit profiles. Deep semantic knowledge provides a framework for extracting rich contextual knowledge of user queries by analyzing the data networks to identify entities in which the users are interested.

The information extraction framework proposed in this paper is consistent with a study [35] which focused on modeling user preferences for customized content recommendation in large knowledge bases primarily relying on data from the Yahoo Knowledge Graph.

III. Methodology

This section focuses on the implementation of information extraction techniques and models for improving knowledge bases based on the SLU framework. Where various knowledge extraction approaches are utilized to identify entities and extract relationships to provide better insights on their application to information extraction based on slot filling and relation detection as the major components of language understanding.

A. *Extracting Information from Personal Knowledge Graphs*

Rapid technological growth over the past few years has caused a drastic increase in the use of smartphones with advanced capabilities in machine learning, speech recognition, virtual assistants, and voice messaging. Spoken Language Understanding (SLU) features in these information gadgets may be used to extract information from knowledge bases through queries, which may be informational, transactional, or navigational depending on the type of operation being performed. Extracting personal information from Knowledge Bases created by smartphone users may require semantic knowledge graphs due to the high likelihood of data variations [41].

This paper uses schema, a Freebase semantic knowledge graph containing 18 different relations concerning the entity people, person, which may be found in a dataset of spoken utterances. For every relation, a complete set of entities extracted from the Freebase knowledge graph are leveraged in querying the specific entity pairs on the internet using the Bing search engine. The SLU semantic space in this work is aligned to Freebase as a back-end semantic knowledge repository to extract knowledge graph relations in the user utterances as illustrated in Fig. 2.

The user utterances are then classified into binary classes, which may be positive or negative depending on their depiction of personal facts. Once the utterances are formulated as a binary classification problem, the Support Vector Machines (SVM) framework is applied to extract refined factual relations. The *SVM^{light}* package is used to classify the utterances implements binary, linear kernels through a one-vs-rest technique [42]. Identifying the entities and their relations in the utterances, a custom personal knowledge graph for that user is populated with the new information, and the process repeats if the user makes further utterances.

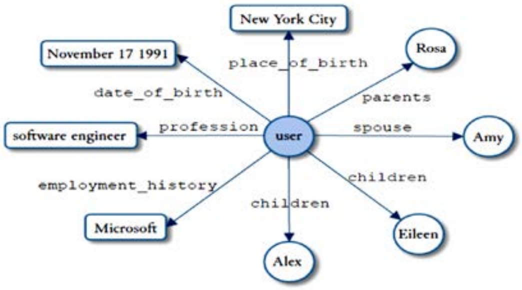

Fig. 2. An Example of a Personal Knowledge Graph.

The training dataset for the framework used in this work is created by searching the internet for related entity pairs in a knowledge graph using the model proposed by [43]. Assuming a web search returns AS as the set containing entity pair a and b, SAS, a subset of AS having

$SAS = \{s : s \in AS \wedge (s, a) \wedge (a, b)\}$ where $\wedge\ (m, n)$is true when n is a substring of string m. The sentences are then post-processed for the augmentation of relation tags from the knowledge base because some instances may contain multiple relations. For example, if two relations; place of birth (New York, USA) and date of birth (October 17, 1983) about “Brad Hudson” is extracted, post-processing would produce the following instances complete with tags instead of tag-less instances: Brad Hudson was born on <date_of_birth>October 17, 1983, </date_of_birth> in <place_of_birth>New York, USA<\place_of_birth>.

B. Classifying the Personal Assertions

In this experiment, 10 million utterances are extracted from Microsoft KBs and query logs. Factual relations are then mined by extracting personal assertions containing factual relations through the following in the following pattern; 'I am a *, I have a * I live * I was born * I work*'. A random subset of the extracted is selected and annotated whether it satisfies the requirements; it is a personal assertion, invokes relations, and entities can be extracted from the invoked relations. The final dataset contains 12,989 personal assertions out of which only 1,811 utterances contain one or more pre-defined relations. A 10-fold cross-validation technique is then used to create 10 random subsamples whereby 9 subsamples are set aside for training and 1 subsample is retained as a validation set. Cross-validation operations are performed only once on each subsample. From the 236,724 collected samples, 234, 650 are classified accurately (99.12%) and 2,074 are classified inaccurately. This implies that SVM is an efficient classifier for personal assertions.

C. Detecting Relations

The performance of relation detection functionality is determined by testing the models trained using the annotated datasets extracted in the previous section in two scenarios; supervised baseline and unsupervised baseline. A precision model *P@N* was used in the evaluation given that *N* represents positive relations in a given set. From the supervised baseline where 2-fold cross-validation is utilized and the model trained on randomly assigned utterances to two data sets, 84.32% *P@N* upper bound precision is obtained while the unsupervised technique attains 42.85% *P@N* upper bound precision.

D. Slot Filling

The supervised technique was used to perform the slot filling operation due to the variations in semantic annotation mechanisms of the sampled set. The slot F-Measure model was applied to the CoNLL processing script, which attained 68.34% performance efficiency. The model achieves higher performance efficiency when applied to minimal annotations and nontrivial tasks as illustrated in Table III.

TABLE III. PERFORMANCE EFFICIENCY RESULTS FOR SLOT FILLING AND RELATION DETECTION IN DATA EXTRACTION

Relation Type	**Count**	**Relation Detection**		**Slot Filling**		
		Unsupervised	***Supervised***	***Supervised***		
		Precision@Count (%)	***Precision@Count (%)***	***Precision (%)***	***Recall (%)***	***F-Measure (%)***
place_of_birth	8	0.00	0.00	0.00	0.00	0.00
religion	8	0.00	50.00	0.00	0.00	0.00
ethnicity	17	0.00	70.59	100.00	17.65	30.00
employment_history	40	7.50	52.50	50.00	12.50	20.00
nationality	47	0.00	63.83	75.00	82.98	78.79
profission	61	0.00	54.10	50.00	1.64	3.72
gender	63	6.35	82.54	90.91	47.62	62.50
date_of_birth	73	46.58	75.34	56.25	36.99	44.63
places_lived	121	2.48	68.59	69.91	65.29	67.52
sibling_s	248	86.29	90.32	85.92	71.08	77.80
children	260	23.08	87.31	80.92	47.31	59.71
parents	401	19.95	86.78	83.97	65.17	73.39
spouse_s	464	82.11	94.39	86.81	68.10	76.33
Total	1811	42.85	84.32	82.01	58.58	68.34

IV. Content Recommendation by User Profiling

The evolution of the Web has positioned the internet as a crucial player in providing users with access to information from multiple sources. Information overload is one of the greatest challenges of the web hence the need for content recommendation to match user interests. Despite the monumental milestones in the design of recommendation systems, there are significant challenges in availing users of high-quality information. This section explores a high-performance content recommendation model for efficient information extraction from knowledge bases. The core objective of user modeling in this framework is to understand their current preferences and predict future interests in contextual applications such as sports databases. The data used in this experiment is obtained from Yahoo News Streams, which contain information such as the sequence of websites that a user has visited as expressed in the (1) for a typical user *u;*

$$\mathbf{L}^{\mathrm{u}} = \langle \mathrm{w}^{\mathrm{u}}, \mathrm{w}^{\mathrm{u}}, \ldots, \mathrm{w}^{\mathrm{u}}, \ldots, \mathrm{w}^{\mathrm{u}} \rangle \tag{1}$$

Such that w^u represents the websites visited by user *u* at time *t*.

Unstructured information containing attributes such as the user location, language, identity, demographics, timestamps, and click/skip labels. Additionally, the Wikipedia Knowledge Graph was used as a knowledge source for enriching feature space by monitoring evolving sources and wrapping different sources.

A. Modeling user Profiles

A high-level Pipeline algorithm is utilized to model user interests and predict preferences and FastEL software is used for linking entities. A separate entity augmentation algorithm is used to extract entities from user logs then link them to the entities in the Wiki KB. The following code is executed to perform this operation;

Input: A sample user opened document *D* stored in a Global KG *G*, which contains relation triples defined by $\sigma = E_{a,}\rho, E_b$ such that ρ represents a relation predicate for *n* iterations *m* maximum augmented entities.

```
Generate initial entities $E = \{e\}$ from $D$
repeat
    Augment entities using facts from $G$
        Re-score interest weights of augmented entities
until converged or reach $n$ iterations
return top $m$ augmented entities from the list
```

Named entities can be extracted from the visited web pages and linked to related Wiki entities based on the user logs. However, the entities may not provide adequate information on user interests hence cannot accurately predict future preferences hence the need to leverage the Yahoo Knowledge Graph to augment the entities into relational facts with a higher degree of accuracy. Once the entities are augmented and retrieved, a decayed interest weight is then assigned to indicate the lowest probability that user interests lie in a particular category.

B. The Framework for Profiling users

According to [44] the user profiling model used for content recommendation in search engines utilizes Factorization Machines (FM) to perform latent factoring and matrix factorization in recommender systems. A latent space for every user is constructed to allow for the differentiation of user preferences in the process of learning from the unstructured dataset. A factorization-machine-based latent factor framework is used to decompose every user profile shared and personalized latent factors. The process of mapping profiles into latent factors is standardized for every user hence making it possible to enrich the information for those with minimal interaction data.

C. Experiment

The experiments are based on a sample of 32.09 billion user logs collected from Yahoo News Stream over one month. The user profiles are evaluated for quality by splitting the dataset into training and testing groups based on event timestamps. Data sets from the first three weeks (23.68 billion events) are used for training while data from the fourth week (8.42 billion events) is used for model testing. For the training dataset, each user profile is ranked and performance evaluated based on ground truth labels, which may be positive or negative.

Inner product values are used between item features and user profiles to generate the ranking scores of each user-item pair. The items are then ranked as positive if they have a higher ranking otherwise negative based on metrics such as the Area under the Curve (AUC), Mean Reciprocal Rank (MRR), and Mean Average Precision (MAP) as defined in the (2), (3) and (4);

$$MAP = 1/m \sum_{i=1}^{m} \frac{\sum_{k=1}^{n_i} P(k)}{n_i} \tag{2}$$

$$MRR = 1/m \sum_{i=1}^{m} \frac{1}{r_i^1} \tag{3}$$

$$MAP = 1/m \sum_{i=1}^{m} \frac{(\sum_j r_i^j) - P_i(P_i+1)/2}{P_i * N_i} \tag{4}$$

Given that *P(k)* represents precision at *k*, n_{i}: user-related links, $u_j r^1$: ranking of the links that were clicked first by user u_i, P_i: the clicked links, and N_i: non-clicked links in the profile for user u_i.

When the number of iterations is adjusted to 1, it achieves about 193% relative and 10% absolute performance improvement in mean average precision; 191% relative and 17% absolute performance improvement in mean reciprocal rank, which is significantly high compared to the baseline system, which obtained 12% relative and 7% absolute performance improvement. Mean average precision computes the average precision scores for listed items while the mean reciprocal rank calculates the inverse position of the initially ranked relevant items. Therefore, both MRR and MAP compute ranking scores for listed items. The area under the curve describes the ratio of false positives and true positives when the threshold parameter is varied suggesting that when entity ranking and coverage are applied to content recommendation through entity augmentation, it extracts additional related entities enriching the feature space significantly according to [45].

V. Discussion and Conclusion

Technological evolution has led to the rapid adoption of online news platforms as a source of information from a wide range of sources across the globe. Due to the high volume of documents on millions of websites, users face many challenges finding their articles of interest or any other precise information. Knowledge Bases such as Wikipedia are rich information resources for users seeking knowledge in various fields including culture, technology, science, and history. This study sought to improve the efficiency of knowledge bases by analyzing the statistical frameworks for building user-centric KBs and extracting personal facts from user utterances through personal assertion classification.

The study also sought to understand how the accuracy of information extracted from knowledge bases can be validated using a maximum entropy framework. Consequently, a framework for rich annotation-guided learning was developed as an approach for improving the efficiency of knowledge basis through information extraction [13]. The annotation framework was designed with a capability for feature enrichment, which allows for the analysis of relative efficacy and scalability of slot filling operations in KBP settings. A review of previously published studies demonstrates that a slight increase in the annotation period improves KB performance significantly. The study also sought to investigate how knowledge bases can be improved to advance tasks such as content recommendation based on the users' online activity. The experimental findings for these improvement operations in knowledge bases suggest that refining information extraction techniques is an efficient approach to improving the performance of knowledge bases.

VI. Future Work

While researchers have made significant progress towards the understanding of knowledge base architectures, various gaps need to be filled, especially on the categories of knowledge possessed by human beings. Current literature does not provide detailed representations of facts based on common sense and procedural knowledge. Knowledge representation through reasoning and learning remains an important aspect of future studies on the integration of machine learning and artificial intelligence capabilities to information extraction from knowledge bases. Other relevant fields for future research include the population of personal knowledge graphs, confidence estimation for knowledge bases, and guided learning for rich annotations.

References

[1] M. Mutasim and A. Karrar, "Impute Missing Values in R Language using IBK Classification Algorithm," International Journal of Engineering Science and Computing, vol. 11, no. 6, pp. 28328-28338, 2021.

[2] J. Ma, D. Li, Y. Chen, Y. Qiao, H. Zhu and X. Zhang, "A Knowledge Graph Entity Disambiguation Method Based on Entity-Relationship Embedding and Graph Structure Embedding," Computational Intelligence and Neuroscience, vol. 2021, pp. 1-11, 2021.

[3] H. Wu, S. Y. Liu, W. Zheng, Y. Yang and H. Gao, "PaintKG: the painting knowledge graph using bilstm-crf," in 2020 International Conference on Information Science and Education , 2020.

[4] T. P. Tanon, G. Weikum and F. Suchanek, "YAGO 4: A Reason-able Knowledge Base," in European Semantic Web Conference, Lecture Notes in Computer Science, 2020.

[5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data," in Lecture Notes in Computer Science, 2007.

[6] D. Diefenbach, M. D. Wilde and S. Alipio, "Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph," in The Semantic Web – ISWC 2021. ISWC 2021. Lecture Notes in Computer Science, 2021.

[7] A. E. Karrar, M. A. Abdalrahman and M. M. Ali, "Applying K-Means Clustering Algorithm to Discover Knowledge from Insurance Dataset Using WEKA Tool," The International Journal Of Engineering And Science, vol. 5, no. 10, pp. 35-39, 2016.

[8] A. E. Karrar, "A Novel Approach for Semi Supervised Clustering Algorithm," International Journal of Advanced Trends in Computer Science and Engineering, vol. 6, no. 1, pp. 1-7, 2017.

[9] N. Sahlab, S. Kamm, T. Müller, N. Jazdi and M. Weyrich, "Knowledge Graphs as Enhancers of Intelligent Digital Twins," in 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems, 2021.

[10] M. Nickel, K. Murphy, V. Tresp and E. Gabrilovich, "A Review of Relational Machine Learning for Knowledge Graphs," Proceedings of the IEEE , vol. 104, no. 1, pp. 11-33, 2016.

[11] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer and C. Welty, "Building Watson: An Overview of the DeepQA Project," AI Magazine, vol. 31, no. 3, pp. 59-79, 2010.

[12] Y.-T. Huang and P.-F. Pai, "Using the Least Squares Support Vector Regression to Forecast Movie Sales with Data from Twitter and Movie Databases," Symmetry, vol. 12, no. 4:625, 2020.

[13] A. E. Karrar, "The Use of Case-based Reasoning in a Knowledge-based (Learning) Software Development Organizations," International Journal of Innovative Research in Science, Engineering and Technology, vol. 5, no. 5, 2016.

[14] A. E. Karrar, N. H. Mohammed and M. M. Ali, "Impact of Using Preprocessing in Data Mining and Knowledge Discovery Process," International Journal of Computing and Technology, vol. 3, no. 12, pp. 524-527, 2016.

[15] S. Qiu, P. An, K. Kang, J. Hu, T. Han and M. Rauterberg, "A Review of Data Gathering Methods for Evaluating Socially Assistive Systems," Sensors, vol. 22, no. 1:82, 2022.

[16] X. Sun, J. Gu and H. Sun, "Research progress of zero-shot learning," Applied Intelligence, vol. 51, no. 2, pp. 1-15, 2021.

[17] T. He, X. Xu, Y. Wu, H. Wang and J. Chen, "Multitask Learning with Knowledge Base for Joint Intent Detection and Slot Filling," Applied Sciences, vol. 11, no. 11, 2021.

[18] A. R. Johansen, C. K. Sønderby, S. K. Sønderby and O. Winther, "Deep Recurrent Conditional Random Field Network for Protein Secondary Prediction," in Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics, Boston, 2017.

[19] F. Orlandi, J. Debattista, I. A. Hassan, C. Conran, M. Latifi, M. Nicholson, F. A. Salim, D. Turner, O. Conlan, D. O'sullivan and J. Tang, "Leveraging Knowledge Graphs of Movies and Their Content for Web-Scale Analysis," in 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2018.

[20] V. Hudeček, O. Dušek and Z. Yu, "Discovering Dialogue Slots with Weak Supervision," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021.

[21] V. Reshadat, M. Hourali and H. Faili, "Confidence measure estimation for Open Information Extraction," Journal of Information Systems and Telecommunication, vol. 6, no. 1, pp. 1-8, 2018.

[22] A. Raghibi and L. Oubdi, "A Proposed Model for Social Impact Sukuk," TURKISH JOURNAL OF ISLAMIC ECONOMICS, vol. 8, no. 2, pp. 501-516, 2021.

[23] L. Qiu, Y. Ding and L. He, "Recurrent Neural Networks with Pre-trained Language Model Embedding for Slot Filling Task," arXiv preprint, arXiv:1812.05199, 2018.

[24] S. Verlinden, K. Zaporojets, J. Deleu, T. Demeester and C. Develder, "Injecting Knowledge Base Information into End-to-End Joint Entity and Relation Extraction and Coreference Resolution," in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021.

[25] C. Orasan and R. Mitkov, "Recent Developments in Natural Language Processing," in The Oxford Handbook of Computational Linguistics 2nd edition, R. Mitkov, Ed., Oxford University Press, 2021.

[26] L. Zhao-Yang and H. Sheng-Jun, "Active Sampling for Open-Set Classification without Initial Annotation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2019.

[27] C. Deng, X. Ji, C. Rainey, J. Zhang and W. Lu, "Integrating Machine Learning with Human Knowledge," iScience, vol. 23, no. 11, pp. 1-27, 2020.

[28] C. David, L. Quentin and D. Alexandre, "Multi-level and multi-scale reconstruction of knowledge dynamics with phylomemies," Scientometrics, vol. 127, pp. 545-575, 2022.

[29] S. Razniewski, A. Yatesa, N. Kassner and G. Weikum, "Language Models As or For Knowledge Bases," arXiv preprint, arXiv:2110.04888, 2021.

[30] A. E. Karrar, "Investigate the Ensemble Model by Intelligence Analysis to Improve the Accuracy of the Classification Data in the Diagnostic and Treatment Interventions for Prostate Cancer," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, pp. 181-188, 2022.

[31] M. Nasir, C. I. Ezeife and A. Gidado, "Improving e-commerce product recommendation using semantic context and sequential historical purchases," Social Network Analysis and Mining volume, vol. 11, no. 1, 2021.

[32] Y. Koren, "Factorization meets the neighborhood: a multifaceted collabora- tive filtering model," in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008.

[33] A. Al-Bazi, V. Palade, R. A. Hadeethi and A. Abbas, "AN IMPROVED FUZZY KNOWLEDGE-BASED MODEL FOR LONG STAY CONTAINER YARDS," Advances In Industrial Engineering And Management, vol. 10, no. 1, pp. 1-9, 2021.

[34] A. Deepak, C. Bee-Chung, G. Rupesh, H. Joshua, H. Qi, I. Anand, K. Sumanth, M. Yiming, S. Pannagadatta, S. Ajit and Z. Liang, "Activity Ranking in LinkedIn Feed," in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014.

[35] E. Zhong, L. Nathan, S. Yue and R. Suju, "Building Discriminative User Profiles for Large-Scale Content Recommendation," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015.

[36] S. Manchanda, M. Sharma and G. Karypis, "Distant-Supervised Slot-Filling for E-Commerce Queries," in 2021 IEEE International Conference on Big Data, 2021.

[37] S. K. Cheekula, P. Kapanipathi, D. Doran, P. Jain and A. Sheth, "Entity Recommendations Using Hierarchical Knowledge Bases," in ESWC 2015, 2015.

[38] M. Umair, F. Majeed, M. Shoaib, M. Q. Saleem, M. S. Adrees, A. E. Karrar, S. Khurram, M. Shafiq and J.-G. Choi, "Main Path Analysis to Filter Unbiased Literature," Intelligent Automation & Soft Computing, vol. 32, no. 2, pp. 1179-1194, 2022.

[39] Z. Fattane, K. Mohsen and B. Ebrahim, "User interest prediction over future unobserved topics on social networks," Information Retrieval Journal, vol. 22, pp. 93-128, 2019.

[40] B. Hui, L. Zhang, X. Zhou, X. Wen and Y. Nian, "Personalized recommendation system based on knowledge embedding and historical behavior," Applied Intelligence, vol. 52, pp. 954-966, 2022.

[41] P. Kaur, P. Nand, S. Naseer, A. A. Gardezi, F. Alassery, H. Hamam, O. Cheikhrouhou and M. Shafiq, "Ontology-Based Semantic Search Framework for Disparate Datasets," Intelligent Automation and Soft Computing, vol. 32, no. 3, pp. 1717-1728, 2022.

[42] S. Nurse and J. Bijak, "Building a Knowledge Base for the Model," in Towards Bayesian Model-Based Demography. Methodos Series (Methodological Prospects in the Social Sciences), vol. 17, Springer, Cham, 2022, pp. 51-70.

[43] W. Wu, Z. Zhu, G. Zhang, S. Kang and P. Liu, "A reasoning enhance network for muti-relation question answering," Applied Intelligence, vol. 51, no. 5, p. 4515–4524, 2021.

[44] S.-Y. Jeong and Y.-K. Kim, "Deep Learning-Based Context-Aware Recommender System Considering Contextual Features," Applied Sciences, vol. 12, no. 1, 2022.

[45] C. Chaudhary, P. Goyal, D. N. Prasad and Y.-P. P. Chen, "Enhancing the Quality of Image Tagging Using a Visio-Textual Knowledge Base," IEEE Transactions on Multimedia, vol. 22, no. 4, pp. 897 - 911, 2020.